Self-replicators from dynamic molecular networks: selection, competition and subsystem coupling

Komáromy, Dávid

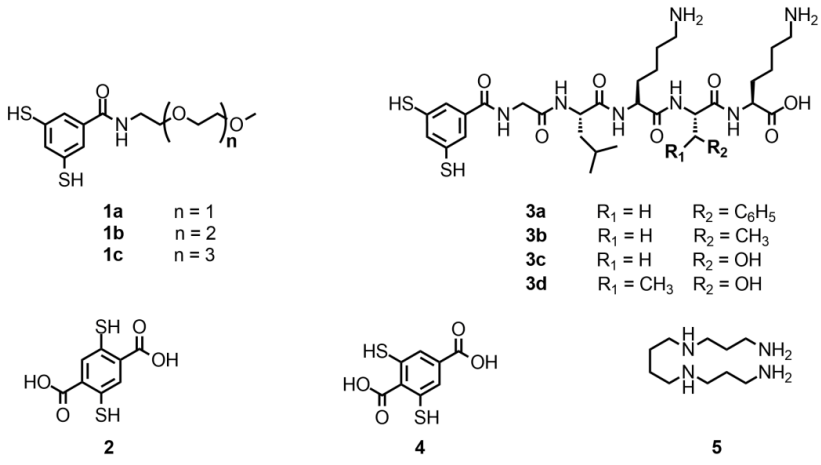
Chapter 6

Summary and Outlook

6.1. Summary

Complex (supra)molecular systems are ubiquitous in living organisms as well as in synthetic context but understanding them in depth is still out of reach, mainly because we do not have much insight into the gradual formation of complex behavior from more simple system elements. The motivation of this thesis is to show how complex chemical systems can be constructed and as a result, how the emergence of novel systems phenomena can be understood with the aid of the applied bottom-up approach. In **Chapter 1**, we introduced the basic concepts which are indispensable for the description of complex chemical systems, such as reaction networks, self-assembly, self-sorting, self-replication, self-organization and subsystem coupling. As a specific type of reaction networks, DCLs (dynamic combinatorial libraries, i.e. interconverting mixtures of molecules appended with functional groups capable of reversible covalent bond formation) were described in detail. Based on our research group's expertise on synthetic chemical self-replicators, we described in depth our well-studied self-replicator systems, based on peptide-dithiol conjugates (enabling dynamic chemistry via formation of thiol-disulfide DCLs), forming macrocyclic disulfides, which auto-catalytically self-assemble into nanoscale fibers, driven by hydrophobic effects between the aromatic cores and structurally well-defined interactions between the peptide chains.

Scheme 6. 1. Chemical structures of building blocks **1**-**4** and guest **5** used in this thesis.



In **Chapter 2**, we showed how non-covalent self-assembly can direct covalent bond formation toward two highly different outcomes, i.e. a diverse set of large macrocycles in one case and one specific macrocycle in the other case. Specifically, a novel oligo(ethylene oxide) dithiol building block **1b** was

described, which, upon oxidation produces two different sets of disulfide macrocycles, depending on mechanical agitation. In the absence of mechanical agitation, unprecedentedly large macrocycles (LMCs) are formed, up to cyclic 44mer. This unusual behavior proceeds via hydrophobicity-driven aggregation of macrocyclic trimers and tetramers, which undergo covalent bond exchange to produce larger species. In contrast, upon mechanical agitation, a self-replicating cyclic hexamer (**1b**)$_6$ (**Scheme 6. 1**) is formed exclusively, self-assembling into nanoribbons. Small changes in the hydrophobicity of the building block structure are translated into large changes at the systems level: building block **1a** gives a self-replicating tetramer and only a minor amount of LMCs in the absence and presence of agitation, respectively, whereas **1c** produces only LMCs regardless of the mechanochemical conditions.

In **Chapter 3** we demonstrated for the first time how self-replication can be triggered by an effector molecule in a system composed of two coupled subsystems. The first subsystem, emerging from building block **1b** is responsible for self-replication, whereas the second one, composed of building block **2** and spermine (**5**) plays the role of effector recognition, with tetramer **2**$_4$ being a nanomolar affinity spermine binder. Upon stirring a solution of **1b** and **2**, a mixed DCL of more than 30 members is formed with no replicator (**1b**)$_6$ present. Upon addition of **5**, the effector-guest complex **2**$_4$·**5** is formed immediately, leaving behind a DCL composed of **1b**-only oligomers. In line with the results from **Chapter 2**, further stirring effected the autocatalytic formation of self-replicator (**1b**)$_6$. This two-step process, besides realizing effector-triggered self-replication, also represents a special case for self-sorting, whereby upon guest addition the originally diverse mixed DCL is transformed into a two-component mixture of (**1b**)$_6$ and **2**$_4$·**5**. Remarkably, the system can be switched several times between the mixed and self-sorted states. Moreover, changing the amount of effector, the onset, rate and extent of self-replication can be tuned. Finally, the approach is modular, i.e. the self-assembly properties of **1a** and **1c** are preserved, when these building blocks are used in the replicator subsystem instead of **1b**.

Having in hand a novel disulfide self-replicator (**1b**)$_6$ with different assembly strength and nanoscale morphology compared to our previously studied peptide-based disulfide self-replicators, in **Chapter 4,** we set out to directly compare the self-assembly properties as well as the auto- and cross catalytic capabilities of the two replicator families. We observed that in DCLs prepared from **1b** and peptide dithiol **3a** (observed to form hexamer replicator (**3a**)$_6$) with increasing amounts of **1b**, mixed hexamers were formed, in molar ratios corresponding to the expected statistical (binomial)distribution. However, at the supramolecular level, no such continuous transition from one assembly form to the other was observed. Instead, the characteristic morphologies of one replicator were no longer observed upon the incorporation of even low amounts (10 mol%) of the other building block. In contrary, at intermediate stoichiometries a novel supramolecular phase emerged, consisting of fibers similar in size to those formed from (**3a**)$_6$, but not featuring supramolecular helicity. The mixed hexamers at 50 mol % **1b** emerged as one set of replicators. Moreover, they were mutually cross-catalytic with pure (**3a**)$_6$. In

contrary, (**1b**)$_6$ only showed catalysis towards its own formation. This trend is ascribed to the similar assembly strength of the mixed replicators and (**3a**)$_6$. Similar behavior was observed for peptide building block **3b** (observed to form octamer replicator (**3b**)$_8$). However, in this case the different macrocycle size for the peptidic and non-peptidic replicators led to remarkable differences at systems level: First, upon gradually increasing the relative amount of **1b**, octamer replicator (**3b**)$_8$ did not persist upon the incorporation of even minor amounts (<10 mol%) of **1b**. Second, the mutual cross-catalysis between pure peptidic and mixed replicators was accompanied by the loss of information arising from macrocycle size, resulting in the emergence of hexameric replicators in all cases. Overall, both macrocycle size and interaction strength between single replicator molecules were observed to determine auto- and cross-catalytic propensities of the different replicators.

Finally, in **Chapter 5**, we demonstrated that in a DCL composed of building blocks **3c** and **2** three distinct disulfide macrocycles featuring entirely different self-assembly properties are formed selectively. Specifically, at very low concentrations of **2**, the self-replicating octamer (**3c**)$_8$ is formed, self-assembling into short nanofibers. At intermediate concentrations of **2** (> 20 mol%), the mixed hexamer (**3c**)$_4$**2**$_2$ is formed, self-assembling into long and thick fiber bundles. At higher concentrations of **2** (> 33 mol%), the mixed tetramer (**3c**)$_1$**2**$_3$ emerges, self-assembling into a [*c3*] daisy chain [(**3c**)$_1$**2**$_3$]$_3$, which becomes the dominant product at higher (≥ 75 mol%) relative concentrations of **2**. The mixed hexamer (**3c**)$_4$**2**$_2$ has been proven to be a self-replicator, although with a remarkably low kinetic barrier of spontaneous formation. In addition to hydrophobic effects organizing the vertical assembly of the macrocycles into fibers, ionic interactions between the negative and positive positively charged lysine ammonium side chains and negatively charged carboxylate moieties are responsible for the lateral association of the fibers into fiber bundles. Similarly, the assembly of the [*c3*] daisy chain [(**3c**)$_1$**2**$_3$]$_3$ is driven by charge complementarity of the aforementioned moieties, as well as by the hydrophobicity of the lysine alkyl chain and the interior of the macrocycle. In other words, the specific self-assembly properties rely on the internal codes for self-assembly programmed into the two building blocks. Moreover, we compared the relative interaction strengths of the different assembly types and found the mixed hexamer to be the thermodynamically most favored assembly due to the multivalent interactions holding together the fibers. Remarkably, both hexamer (**3c**)$_4$**2**$_2$ and tetramer daisy chain [(**3c**)$_1$**2**$_3$]$_3$ are formed exclusively at stoichiometries corresponding to their composition (33 and 75 mol% **2**, respectively). Whereas in principle a high number of other DCL members could have formed even at biased stoichiometries, the above finding points out the strong and specific secondary interactions which organize the assemblies. We additionally investigated the modularity of the approach and found that closely related peptide analogues **3b** and **3d** both give rise to the same systems behavior, whereas in the case of **3a** only the mixed tetramer daisy chain emerges specifically, and no macrocycle is preferentially formed by combining **3c** and non-peptide building block analogue **4**.

6.2. Conclusions and Outlook

After years of research in systems chemistry, we are particularly called to reflect on the broader implications of our findings: specifically, what problems do they solve, what do we learn from them and what opportunities do they create from the systems chemist's perspective. First, as we already indicated in Section 1.5., our findings did not result from the attempt of resolving previously unresolved problems (as it would be the case, were we given a research question of a different kind, e.g. that of elucidating a certain reaction mechanism or developing an important but missing analytical method). They were rather the result of a number of serendipitous findings, which we later tried to explain and tried to place in the context of supramolecular systems chemistry. Consequently, it would be rather unscientific (and untrue to the nature of scientific discovery, as indicated in Section 1.5.) to try to invent a research question *afterwards*, to which our findings would seem to give an answer. This rather unconventional situation arises from the nature of systems chemistry, as the topic of this discipline is as much the creation of novel phenomena as the explanation of already known ones. Thus, we leave the question concerning the *problems* solved in this thesis open. However, we firmly believe that there are certain general *trends* that can be abstracted from our results; furthermore, that these results open up the field for new *opportunities* in systems chemistry and we would like to outline those here.

6.2.1. Conclusions

In general, the purpose of this study is the design of chemical systems which can fulfil certain function(s) not possessed by their constituent molecules. During our investigations, we focused on the development of such systems arising from the interaction of multiple self-assembling (mostly self-replicating) subsystems. In order to develop systems with diverse functionalities, first, independent elements (or subsystems) with well-defined characteristics have to be constructed (see Section 1.4.); second, these elements must be interconnected along (a) well-defined pathway(s). In this section, we summarize the most important trends and challenges related to the goal stated above.

In our case the mutual independence of the subsystems mostly relies on orthogonal self-assembly processes, i.e. on self-sorting. From **Chapter 3** it is clear that the secondary interactions enabling these self-assembly processes must be rather strong, specific and possess different kinetic and/or thermodynamic parameters. Otherwise, as shown in **Chapter 4**, the boundaries of the individual subsystems disappear and no self-sorting takes place.

A related problem involves the search for novel secondary interactions that enable self-assembly, which can occur even in the presence of other self-assembling species. As seen in **Chapter 5**, such a secondary interaction (or binding motif) can be rather specific. In this regard, the systems chemist might

make use of the numerous different donor-acceptor pairs (e.g. radical-radical, CH-anion etc.) utilized in classical supramolecular chemistry to construct novel types of self-assembly.

As for self-replication, we clearly saw in **Chapter 2** that non-directional, hydrophobicity-driven secondary interactions are sufficient to induce self-replication, i.e. one does not need a specific hydrogen bonding pattern or peptide sequence as a structural element to induce self-replication. Consequently, when designing e.g. functional self-replicators, one might first equip a certain building block with chemical moieties that expectedly give rise to a certain functionality; in the second step of the design, the hydrophobicity (and hence, the self-replication ability) might be tuned by linking chemically inert chemical moieties (e.g. oligo(ethylene oxide) chains) to the building block designed in the first step. In other words, our finding might render the design of self-replicating functional molecules more modular.

Finally, an important step in the development of self-replicating systems is related to the problem of information transfer. If we attempt to construct chemical systems (be they life-like or not) which can store and transmit larger amounts of data, it should be able to self-replicate sequence- specifically. Our results from **Chapter 5** show that this requirement might be approached by introducing strong and specific secondary interactions into the constituent building blocks. However, our results are preliminary in this regard and more examples are needed in order to determine the requirements for sequence-selectivity. Furthermore, regarding the difficulties of structure elucidation described in the same chapter, our analytical methods should be improved in order to allow for more precise sequencing of the replicators.

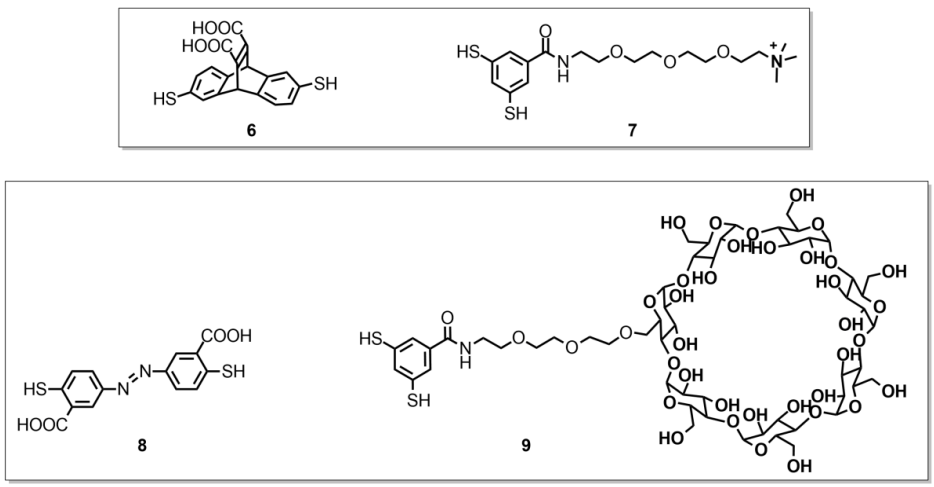
6.2.2. Outlook

After the most important conclusions from our work, we attempt to sketch a few research topics based on the trends outlined above.

First, subcomponent self-assembly of cyclic daisy chains in water as presented in **Chapter 5** would be worthwhile to be studied in depth. For example, as not every amino acid moiety seems to be necessary for the formation of the interlocked structure, the minimal required peptide motif for daisy chain formation has yet to be uncovered. Different peptide sequences might lead to daisy chains with different numbers of tetramer units. Furthermore, using multiple peptide building blocks simultaneously (e.g. a mixture of **3a**, **3c** and **2** in a molar ratio of 1:1:6) might effectuate self-sorting at the molecular level (i.e. selective formation of $\mathbf{(3a)_1 2_3}$ and $\mathbf{(3c)_1 2_3}$), but not necessarily at the supramolecular level (i.e. statistical formation of mixed $\mathbf{[(3a)_1 2_3]_n[(3c)_1 2_3]_{3-n}}$ daisy chains). Finally, from the perspective of *de novo* life research, the use of the novel daisy chains as artificial molecular muscles can be considered, i.e. covalent capture at the lysine ammonium groups and subsequent study of the internal motion in the so-formed cyclic rotaxane-based daisy chains upon protonation and deprotonation.

From a more general point of view, the findings in **Chapter 5** might inspire the exploration of multicomponent disulfide DCLs constructed from building blocks featuring complementary donor-acceptor motifs. For example, it is known that the tetrameric disulfide formed from building block **6** is a strong acceptor for tetramethylammonium ions.[1] Thus, a DCL composed of **6** and putative building block **7**, bearing a tetramethylammonium moiety, might display similar behavior to the system described in **Chapter 5** (note that the charge properties are also similar to that of the previously studied DCL). On a different note, dithiol **8** has been reported to form macrocyclic trimers and tetramers, both of which form 1:1 catenanes with β-cyclodextrin.[2] Thus, mixing **8** with putative β-cyclodextrin-conjugated dithiol **9** might give rise to even multiple distinct interlocked complexes which might form selectively at given stoichiometries.

Scheme 6. 2. Reported (**6**, **8**) and designed (**7, 9**) building blocks expected to display selective formation of mixed macrocycles and/or interlocked structures based on the presence of complementary donor-acceptor motifs in the building blocks.



On the way towards the construction of interconnected subsystems (**Chapter 3**) one might realize that a higher degree of separation of the subsystems is required. This requirement might be brought about by using compartmentalized subsystems, and/or using different (possibly orthogonal) dynamic combinatorial chemistries for the different subsystems. Notably, this attempt will likely require careful optimization: although orthogonal DCCs have been reported,[3] their study has not substantially went beyond the mere demonstration of their orthogonality. In other words, the possibility of coupling of alternative dynamic covalent groups to moieties which can bring about self-assembly is still awaiting exploration.

6.2.3. Systems Chemistry and the Current Culture of Science

In this thesis we described several novel self-replicators based on disulfide chemistry. However, years of research in DCC and systems chemistry showed (at least to the author of this thesis) that the discovery of novel self-replicating systems might be painfully slow and highly governed by serendipity if conventional, rational design and manual synthesis is applied. Additionally, in the meanwhile, a huge number of "negative" findings are described (i.e. building blocks which were expected to form self-replicators but failed to do so), which are nevertheless not published. These problems might be circumvented (and even made use of) by the publication of negative results and the use of artificial intelligence (AI) on the obtained data sets. In the recent years, AI has been proven successful in various fields of chemical discovery,[4] e.g. in finding novel organic reaction mechanisms,[5] medicinally relevant compounds[6] or novel inorganic materials, using failed experiments.[7] Taking into account the modular synthesis of peptide-based replicators, an approach combining automated synthesis of putative building blocks, microfluidics-based production and analysis of DCLs and AI-based evaluation of the results, as well as a semi-automated feedback loop towards the structural design of novel candidates might greatly facilitate the discovery of novel self-replicators. Notably, this approach would not "take the jobs" of systems chemists, rather shorten the time invested on potentially unfruitful work.

In the light of the suggestions described above, one might radically change one's perception on success and failure in the context of scientific research. First, by making use of failed experiments by AI and machine learning, the value of negative results might substantially increase and thus might not be dismissed any longer from the scientific literature (One should note that this perspective change does not necessarily require high-throughput computational methods as proven by e.g. the existence of the *Journal of Negative Results in Biomedicine*).[8] Second, this approach might positively contribute to the overall mental state of researchers who until now have been forced to present positive results, which often come to light slowly, painfully and unexpectedly.

6.3. References

[1] P. T. Corbett, L. H. Tong, J. K. M. Sanders, S. Otto, *J. Am. Chem. Soc.* **2005**, *127*, 8902–8903.

[2] J. Li, P. Nowak, H. Fanlo-Virgós, S. Otto, *Chem. Sci.* **2014**, *5*, 4968–4974.

[3] H. M. Seifert, K. Ramirez Trejo, E. V. Anslyn, *J. Am. Chem. Soc.* **2016**, *138*, 10916–10924.

[4] A. Aspuru-Guzik, R. Lindh, M. Reiher, *ACS Cent. Sci.* **2018**, *4*, 144–152.

[5] J. M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, *Nature* **2018**, *559*, 377–381.

[6] H. Altae-Tran, B. Ramsundar, A. S. Pappu, V. Pande, *ACS Cent. Sci.* **2017**, *3*, 283–293.

[7] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73–76.

[8] R. S. Hebert, S. M. Wright, R. S. Dittus, T. A. Elasy, *J. Negat. Results Biomed.* **2002**, *1*, 1.